DOCUMENT RESUME

ED 076 564 SP 006 495

AUTHOR Siegel, Martin A.; Rosenshine, Barak
TITLE Teacher Behavior and Student Achievement in the Bereiter-Engelmann Follow-Through Program.
PUB DATE Feb 73
NOTE 33p.; Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, La., February 1973

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Academic Achievement; Instructional Materials; Research Projects; *Response Mode; Student Teacher Relationship; *Teacher Behavior; Teacher Evaluation; *Teaching Techniques

ABSTRACT

The purpose of this study was to determine the relationship between teacher behavior and student achievement in the Bereiter-Engelmann program. Ten groups were observed in the first study, 24 groups in the second. All teachers were rated on four occasions using a highly specific rating scale. The pre- and postmeasures were criterion-referenced. Four variables remained in predictive importance across studies: following the format, requiring 100 percent criterion responding, correcting mistakes, and presenting signals. Since the most critical variables affecting student gains may be those which are not included in general observational instruments, development of instruments specific to a curriculum program seems useful. (Author)

Introduction

Among the many curriculum materials packages that have been developed, there has been little research that has investigated the relationship between teacher behaviors as prescribed by the curriculum developers and student outcomes such as achievement or attitudes. The research on teacher behaviors within curriculum packages generally falls into two major categories: 1) studies which describe curriculum relevant teacher behaviors but do not relate these activities to student growth (e.g., Olivero, undated; Gallagher, 1966, 1968; Katz, 1968; Lindvall and Cox, 1970; Niedermeyer and Dalrymple, 1970; Bissel, 1971) and 2) studies relating general instructional activities to student outcomes (e.g., La Shier, 1967; Walberg, 1969; Flanders, 1970; Soar, 1971; Soar, Soar, and Ragosta, 1971).

Unfortunately the results of especially the first group of studies can have limited impact on the development or assessment of the teacher training programs within specific curriculum packages, or on the modification of the curriculum materials themselves. The descriptive studies, although suggesting wide variation in events within classrooms using a particular curriculum package, do not relate the variation to student outcome measures. For example, Gallagher (1966) counted various types of activities which occurred in the classrooms of six teachers who were teaching the same unit from the Biological Sciences Curriculum Study (BSCS) program. On almost all measures of teacher behavior there were significant differences among the six teachers. Regrettably, the investigator did not relate this variation to measures of student outcomes. Does an increase in inquiry-strategy behaviors which are intended by the BSCS curriculum planners enhance or suppress student achievement or is the effect negligible? Given a behavior that affects cognitive gains, what are the concommitant effects in attitude towards the curriculum, towards the school, or towards the child?

Introduction

Among the many curriculum materials packages that have been developed, there has been little research that has investigated the relationship between teacher behaviors as prescribed by the curriculum developers and student outcomes such as achievement or attitudes. The research on teacher behaviors within curriculum packages generally falls into two major categories: 1) studies which describe curriculum relevant teacher behaviors but do not relate these activities to student growth (e.g., Olivero, undated; Gallagher, 1966, 1968; Katz, 1968; Lindvall and Cox, 1970; Niedermeyer and Dalrymple, 1970; Bissel, 1971) and 2) studies relating general instructional activities to student outcomes (e.g., La Shier, 1967; Walberg, 1969; Flanders, 1970; Soar, 1971; Soar, Soar, and Ragosta, 1971).

Unfortunately the results of especially the first group of studies can have limited impact on the development or assessment of the teacher training programs within specific curriculum packages, or on the modification of the curriculum materials themselves. The descriptive studies, although suggesting wide variation in events within classrooms using a particular curriculum package, do not relate the variation to student outcome measures. For example, Gallagher (1966) counted various types of activities which occurred in the classrooms of six teachers who were teaching the same unit from the Biological Sciences Curriculum Study (BSCS) program. On almost all measures of teacher behavior there were significant differences among the six teachers. Regrettably, the investigator did not relate this variation to measures of student outcomes. Does an increase in inquiry-strategy behaviors which are intended by the BSCS curriculum planners enhance or suppress student achievement or is the effect negligible? Given a behavior that affects cognitive gains, what are the concommitant effects in attitude towards the curriculum, towards the school, or towards the child?

While the second group of studies do attempt to relate instructional activities to measures of student outcomes, the observational instruments used were designed to apply to all types of programs and educational settings. For example, Soar (1971; Soar, Soar, and Ragosta, 1971) has been monitoring eight classrooms in each of seven Follow-Through programs along with two comparison classrooms for each program. Instead of developing program-specific observation instruments Soar used four general observational systems: the Reciprocal Category System (Ober, #61[1], an expansion of the Flanders system, #5), the Florida Taxonomy of Cognitive Behaviors (K-1 Form) (Brown et al., #37), the Teacher Practices Observation Record (Brown, #36), and the Florida Climate and Control System (Soar, 1966; Soar, Soar, and Ragosta, 1971). Although the investigators correlated the factor scores derived from the four instruments with measures of class mean residual gain, it is conceivable that the most critical variables which affect student gains are those which were not included in the general observational instruments. The ability to follow a pre-specified format without even minor deviations may be an important variable in the Engelmann-Becker program, whereas in the Bank Street Program, the ability to elaborate on a child's experiences may be essential to the realization of the programs goals and objectives. However, a general observation instrument is likely to be insensitive to either of these program-specific variables. Therefore, in addition to general instruments, development of observational measures specific to the instructional activities most emphasized by the curriculum designers seems useful.

---

[1]Numbers such as this refer to those assigned each observational system in Mirrors for Behavior (Simon and Boyer, 1967, 1970a, 1970b).

Rosenshine and Furst (1972) suggested that research on a particular curriculum materials package should consist of five phases:

1. Train a group of teachers to use a certain package of materials which have already received extensive trial and modification within special settings (for example, any of the Follow-Through programs like the Bank Street Program, Bushell's Behavior Analysis Program, or Engelmann and Becker's Distar Program; BSCS; First Year Communication Skills Program; or Harvard Project Physics).

2. Use observational systems to describe instructional variables which are considered specific to the program and most emphasized by the curriculum planners and which are also considered to have general educational importance (and may or may not be emphasized by the curriculum designers).

3. Study the relationship between instructional activities and behavioral change in the students in a variety of outcomes. At least the following ten questions should be asked (Rosenshine, 1971):

    1. To what extent were the instructional activities within the program those which were intended by the curriculum developers?

    2. Did the classrooms (or other units) within the program differ in their use of instructional activities specific to the program?

    3. Did the classrooms within the program differ in their use of general instructional activities considered important for student growth?

    4. Were the classrooms within the program different on the outcome measures of interest?

    5. What was the relationship between use of program-specific activities and student growth?

    6. What was the relationship between general instructional activities and student growth?

    7. Were there differences in student growth among classrooms of teachers who were high, average, or below average in their fidelity to the intentions of the curriculum developers?

8. Were there differences in student growth among classrooms of teachers who were high, average, or below average in their use of general instructional activities?

9. Were classrooms which were high, average, or below average in student growth different in their fidelity to the curriculum developers?

10. Were classrooms which were high, average, or below ~~average in student growth~~ different in their use of general instructional activities?

Rosenshine in a later paper (1972b) argues against averaging implementation ratings across visits for those teachers whose ratings increase, decrease, or are erratic throughout the year (e.g., low, average, high; high, average, low; or high, low, average). Averaging ratings and describing the classrooms as medium implementors is not particularly indicative of what happened. Rosenshine would therefore add to his implementation categories in questions (7) and (8), depending on the meaningful patterns that emerge, the categories of "ascendant," "descendant," and/or "erratic." These patterns could also occur in student behavior (if measures of student outcomes were taken at different intervals throughout the year) and therefore the categories in questions (9) and (10) would be increased.

Problems and suggestions for selecting measures of instructional behaviors and student growth on outcomes of interest, and for data analysis and design are presented elsewhere (Medley and Mitzel, 1963; Gage, 1969; Flanders, 1970; Rosenshine, 1970a, 1970b, 1971; Rosenshine and Furst, 1971, 1972; Tatsuoka, 1972).

4. Modify the training procedures and/or materials on the basis of the studies completed in phases two and three.

5. Conduct new studies with appropriate control groups to determine the effects of the modifications and to determine the new relationships between instructional activities and student growth. By recycling through phases

one through four, the curriculum designer, publisher, and researcher successively approximate optimum training procedures, thus affecting gains in student achievement or other measures of interest.

In a paper on nationwide evaluation and experimental design, Tatsuoka (1972) suggested an evaluation procedure similar to Rosenshine and Furst's "descriptive-correlational-experimental-loop." Tatsuoka emphasized the necessity for random assignment of all units (classes, schools, or school districts) to treatment and control conditions. He, furthermore, defines "experimental treatment" within educational research as an everchanging entity.

> ...in a laboratory experiment in which the treatments are completely specified a priori--such as fixed dosages of a drug, or certain methods of stimulus presentation--these must be held constant throughout. But an educational program is, by its very nature, an entity that is in perpetual flux. Only some broad guidelines and principles are typically specified at the outset, and details of how to carry out the program are usually left to the individual administrator to plan and modify with experience. This fluid, dynamic entity, with all its periodic modifications and refinements IS the treatment. Nothing in experimental design forbids such types of treatment. All that is required is that an accurate running record be kept of what sorts of modifications and refinements were made at what stage for what reasons, so that upon completion of the evaluation we can describe what it is that has been evaluated (p. 3).

Although Tatsuoka's and Rosenshine and Furst's design (developed independently) for curriculum research and evaluation is not particularly unique, no studies were found which included all phases of the design. Research studies which include part of the"loop" exist. However, even this type of instructional research within curriculum programs is rare. In fact, only two studies were found which included the training, descriptive, and correlational phases and also used program-specific variables: Kochendorfer (1967) and Baker (1969). One study (Rosenshine, 1972a) was found where the author reanalyzed the data from a report on the First Year Communication Skills Program (Resta and Hanson, 1971) to obrain correlations between the eleven transactional program-specific variables and class residual gain scores.

Kochendorfer (1967) monitored the practices of 64 biology teachers. These practices were determined by use of a Biology Classroom Checklist, completed by students in one of each teacher's classes. This instrument was developed by Kochendorfer to determine the extent to which BSCS and non-BSCS teachers were using classroom practices recommended by BSCS. The Processes of Science Tests (developed by BSCS staff) was given to detect changes in student understanding of science--to interpret data and deal with hypotheses. The teachers completed an Attitude Inventory (Blankenship, 1965) as a measure of their acceptance of the published BSCS philosophy and rationale. Significant ($p < .01$) differences were found in the classroom practices of experienced BSCS, first-year BSCS, and non-BSCS teachers. A significant ($p < .02$) relationship between the nature of the classroom practices and gains of the Processes of Sciences Test was found (.32). A significant ($p < .02$) correlation was also found between the teacher's attitude concerning the BSCS philosophy and rationale and the degree to which his classroom practices agreed with those advocated by BSCS (.73).

In another study, Baker (1969) trained 38 Peace Corps trainees in the use of theoretically-based learning principles. Observers were concurrently trained to record teachers' use of these principles. The trainees were then required to teach high school students in a videotaped lesson. Trainees were each assigned a behavioral objective to achieve and high school students were pre- and posttested on items measuring the objectives. Even in the somewhat restricted range of behavior, significant ($p < .05$) positive relationships were found between student achievement and trainees' ovserved use of the principles of "appropriate practice," (.34), "individual differentiation," (.43) and "knowledge of results," (.31).

The overall objective of undertaking the following two studies was to partially verify and illustrate the above methodology for determining those features of programs and teacher behavior which appear to be crucial in enhancing student growth. Only the first three phases were undertaken in the studies presented in this paper. A larger study is currently in progress which includes all phases of the research.

Study One

The purpose of the following study was to determine the relationship between those teacher behaviors emphasized in training and measures of student achievement in the Distar Instructional System, a highly structured curriculum materials package.

Method

Description of the Distar Instructional System. Perhaps one of the most successful and controversial of all the early childhood curriculum materials programs is Distar Reading, Language, and Arithmetic (Engelmann and Bruner, 1969, 1970; Engelmann and Carnine, 1969, 1970, 1972; Engelmann, Osborn, and Engelmann, 1969, Engelmann and Osborn, 1970, 1972; Engelmann and Sterns, 1972), a commercial model of the Engelmann-Becker (Bereiter-Engelmann) Follow Through program.[2] Unlike other programmed materials, the Distar program is not a self-instructional program. Instead, the teacher follows a carefully structured and logically sequenced teaching program. The presentation books provide the teacher with a script, a series of demonstrations and tasks to be presented word for word. The teacher's role thus changes from one of designing instruction to one of teaching a particular format to criterion, involving all of the children in the instruction, correcting mistakes, providing feedback, and reinforcing the children's responses.

A thirty-minute lesson consists of a series of tasks that the teacher presents from the presentation book. The tasks consist of group and individual activities. Once the teacher obtains the children's attention she proceeds with the first task, following the format as written in the presentation book. The students respond and the teacher evaluates their answers. If the responses are appropriate, she provides praise or other forms of reinforcement. If, however, one of the children answer

---

[2] For a more complete outline of the philosophy and methods used in the Engelmann-Becker program the reader is referred to Engelmann (1969a, 1969b) and Maccoby and Zellner (1970).

incorrectly, the teacher corrects the mistakes according to a pre-specified correction paradigm. After all of the tasks in the lesson have been presented in this manner, the teacher presents reinforcement material from the "take homes." Then she awards the "take-homes" to children who have performed well. During the next session, the class moves on to the tasks in the following lessons.

An example of a format appears in Figure 1. This task is the first task in Distar Reading I; its purpose, along with other formats, is to teach the skill of sequencing.

Basic Teaching Assumptions. Since the first Distar program was published by Science Research Associates in 1969, certain basic assumptions as to how the teacher should behave when implementing the curriculum materials have been stated explicitly. Five areas of teacher behavior are emphasized throughout teacher guides and training manuals:

I. Following the Format

> The pictures and tasks in the Distar Program are not designed to provide you with points of departure for discussions. They are designed to achieve very specific objectives. These objectives will not be met if you talk too much, if you allow the children to make too many extraneous observations, or if you depart from the task as it is specified in the program.
>
> Use the exact wording provided in the materials, and do not make additional statements or ask additional questions unless the format calls for them. Let the children know that you are on the task. Discourage irrelevant observations. (Distar Language II Teacher's Guide, p. 12)

II. Signals

> Use clear signals for the children to respond, so that they all respond at the same time. The children aren't performing acceptably unless all of them respond appropriately to every question. If some do not respond to a question, the group's response is unacceptable. In such a situation, some children may be learning to listen to what others say and imitate their responses. . . . With clear signals, you will be able to get much more accurate feedback from the performance of the different children in the group (Distar Arithmetic III Teacher's Guide, p. 14)

1 START THE PROGRAM HERE.

Task

a. Everybody, look at me! Praise the children who look.

b. Clap your hands once; pause; slap your lap once with both hands.
This is the right way.

c. Do the sequence four times. Go slowly.
Before each sequence, say: Again.
After each sequence, say: I did it the right way.

d. Have the children do the sequence with you eight times. Go slowly.
Do it with me.
Before each sequence, say: Again.
After each correct sequence, say: We did it the right way.

e. Give each child a turn. Let's see you do it the right way.
Praise the children for correct responses.
· To correct: Repeat d. If a child does not do it correctly after several tries, praise him for trying and go to another child.

f. Everybody, look at me! Praise the children who look.
Tap your head once; pause; stamp your foot once.
Is this the right way? Wait. Praise the response "no."
Show me the right way.
· To correct: Let's do it the right way.
Repeat the correct sequence and then f.

g. Everybody, look at me! Praise the children who look.
Clap your hands once; pause; slap your lap once with both hands.
Is this the right way? Wait. Praise the response "yes."
· To correct: Let's do it the right way.
Repeat the correct sequence and then g.

h. Everybody, look at me! Praise the children who look.
Slap your lap once with both hands; pause; clap your hands once.
Is this the right way? Wait. Praise the response "no."
Show me the right way.
· To correct: Let's do it the right way.
Repeat the correct sequence and then h.

NOW GO TO BLENDING--SAY IT FAST

Figure 1. Example of a format in the Distar Reading I program. What the teacher says is underscored.

This format and other formats in this paper are copied with the permission of Science Research Associates, Chicago.

III. Corrections and Requiring 100% Criterion Responding

Correct only the part of the exercise the child had trouble with. Correct the mistake immediately after it occurs.

After correcting the child on the part of the task he missed, always return to the beginning of the exercise and repeat the exercise. The reason for this procedure is that the children must learn to see each exercise as a series of steps. The steps do not occur in isolation. They are related to a goal and to certain rules. Unless you always repeat a task from the beginning and do not conclude that the children have been corrected until they can go through the entire exercise without making a mistake, the children may learn to handle each of the steps without ever seeing how the steps fit together in a pattern.

Remember--after every mistake return to the beginning of the task and take the entire group of children (not merely the child who made a mistake) through the exercise from the beginning, either until the children are firm or until they make their next mistake (at which time you correct and then return to the beginning of the task). (Distar Arithmetic III Teacher's Guide, p. 14)

IV. Praise and Feedback

Reinforce the children who are on task. Follow the rule of catching children in the act of being good. Show the misbehaving child that he is receiving no rewards and that the children who are working are receiving rewards. (Distar Language II Teacher's Guide, p.14)

Always relate the performance of the children to the rules. Do so in a positive manner. . . . Give the children feedback on each of the behaviors that enter into working hard. This means that you should let the children know when they are working hard. 'Working hard' actually covers a variety of behaviors; giving the correct response; following your presentation--looking at the chalkboard, listening and responding to instructions, answering questions. (Distar Training Level I Participant's Manual, p. 60)

V. Pacing

Pace your presentations so that you move quickly in the right places but slowly when necessary. Move rapidly enough for the children to see the point of each task--always at a rate that will maintain their interest and enthusiasm. (Distar Language II Teacher's Guide, p. 14)

These are the basic implementation variables. It is assumed that if a teacher behaves in these ways the children will achieve the academic objectives of the Distar program. That is, the above teacher behaviors are directly related to student achievement. The following quotation from the introduction to the Distar two-day orientation-training manual indicates this belief:

> You (the teacher) should learn how to present the tasks so that even the lowest-performing children will learn rapidly. Without this workshop training, the chances are that you will not teach the lowest performers in your class. With the training, however, you should be able to reach children that you have not been able to reach in the past. The teaching techniques that you practice here will help you become a better teacher of all your children, but will make the biggest difference with your low-performing children. (p. 2)

Procedure. Ten teachers were given a one-week orientation workshop in which they were trained to teach the Distar Language I program. Extensive amounts of time were spent in training the teachers to follow the formats, the techniques for correcting mistakes, and the principles of behavior modification.[3]

At approximately lesson 47, the lowest performing group of each teacher (each group consisting of five to eight kindergarten or first grade children) was verbally tested on an 84 question criterion-referenced test covering lessons 1 through 80 of the program. In general, the test was constructed in such a way that none of the exact questions asked in the program was used on the test. For example, if the program had a picture of a ball over a table and the question in the program was: "Where is the ball?", then a question on the test might be of the same form but the picture would be different (e.g., a picture of a shoe over a banana with the question, "Where is the shoe?"). The test included the following concepts:

---

[3]The training manual used during the workshop was Distar Orientation: Participant's Manual. This manual is published by Science Research Associates, 1970.

identity statements: When asked questions pertaining to the identity of common objects the child should answer in complete affirmative and negative statements.

polars: When asked to identify and use descriptive adjectives and their opposites (polars), the child should formulate complete affirmative and negative statements.

prepositions: When asked to answer questions about the location of objects, the child should answer in complete statements containing prepositions.

categories: When asked to identify and classify objects, the child should apply rules of classification to the objects to determine whether they fit within a given category.

plurals: When asked questions about the identity of singular and plural objects, the child should produce complete affirmative and negative statements.

parts: When shown an object, the child should identify it, distinguish the parts from the whole, name the object's parts, and give the function of the object and its parts.

Because of various scheduling complications, the tests were not administered to the children on exactly the same lesson in the program. That is, one child may have received the pretest on lesson 36, and another child on lesson 49. Nevertheless, each child received the posttest (identical to the pretest) when he reached lesson 80 in the program.

Four lessons were chosen at random for each teacher, and the entire thirty-minute lesson was audiotaped for analysis. The teacher did not know that she was to be taped until about ten minutes prior to teaching the lesson. Each teacher was rated on a five point scale as follows: 5 = Excellent; 4 = Very Good; 3 = Good; 2 = Adequate; and 1 = Not Acceptable.

Three graduate students rated all recordings while following the exact script the teacher was to follow for that particular lesson. The variables selected for observation and analysis were those most stressed in the training program and considered most important for the success of the Distar program.

Observation Instrument. The eight categories upon which each teacher was rated and the criteria for receiving an excellent rating were as follows:

1. Follows the Format for Group Tasks

   Says all the words in the teacher's script, not omitting words and not adding words except to praise, correct, or require 100% criterion responding.

   Never leads the children (responding with them) or gives spurious cues when they are to respond.

2. Follows the Format for Individual Tasks

   (Same as "Follows the Format for Group Tasks")

3. Corrections

   Corrects all mistakes as they occur according to the following paradigm:

   A. If the child understands the signal but lacks information:
      1. Teacher gives the answer or the least amount of information needed to correct the child's error.
      2. Teacher tests the child by repeating the segment of the task that was missed.
      3. If the mistake occurred in the middle of a task with more than one segment, teacher then repeats the entire task from the beginning.

   B. If the child understands the signal but lacks the motor ability to produce the response:
      1. Teacher leads the child; that is, he repeats the response with the child several times.
      2. Teacher tests the child by repeating the segment of the task that was missed.

   C. If the child does not understand the signal:
      1. Teacher repeats the signal.
      2. Teacher or another child models the segment that was missed.
      3. Teacher tests the child by repeating the segment that was missed.
      4. If the mistake occurred in the middle of a task with more than one segment, teacher then repeats the entire task from the beginning.

4. Requires 100% Criterion Responding

   Requires correct observable responses from all the children to the signals which have been established in the task.

Brings the group to 100% mastery on all parts of the task before continuing with the next task.

Returns to the beginning of the task after correcting a segment of the task.

5. Signals

Pauses before signals.
Presents clear signals--they are "followable."

6. Praise and Feedback

Praises the children for appropriate responding and attending behavior.
Often repeats the correct response.
Relates all praise to the signals established in the task.

7. Pacing Within Tasks

Moves quickly after getting the children's attention.
Chains the parts of a complex task together.
Changes inflections and talks at different levels of loudness.

8. Pacing Between Tasks

Does not waste time between tasks.
Is not sidetracted by children's comments which do not pertain to the task.
Acts generally unpredictable between tasks.
Does not spend a great deal of time reinforcing the children.

Results

Correlations between the ratings on the eight categories and student achievement (adjusted by regression for the mean pretest score and the mean lesson number when the pretest was given) are reported in Table 1. All categories showed a significant ($\alpha = .05$) correlation. Inter-rater reliability for each of the categories ranges from .85 to .97.

Examination of the scatterplots, however, shoed that two of the ten teachers had classes which achieved much less than the other eight teachers, and these two teachers also received markedly lower ratings. Thus, the unbelievable

Table 1

Correlations for Study 1[a]

	Means	Standard Deviations	Part Correlations	T-Score Part[b] Correlations	Partial Correlations
Pretest Score	59.03	6.66			
Day in Program	47.82	7.15			
Posttest Score	73.81	8.70			
1. Follows the Format Group Tasks	3.98	1.32	.77***	.72**	.95***
2. Follows the Format Individaul Tasks	3.80	1.31	.74**	.69*	.83***
3. Corrections	3.50	1.16	.86***	.83***	.96***
4. Requires 100% Criterion Responding	3.90	0.98	.89***	.85***	.99***
5. Signals	3.30	1.06	.83***	.77***	.87***
6. Praise and Feedback	3.90	0.66	.68*	.32	.71*
7. Pacing Within Tasks	3.40	0.70	.84***	.54	.86***
8. Pacing Between Tasks	4.25	0.94	.77***	.46	.86***

*** $p < .005$ (one-tail)

** $p < .01$ (one-tail)

* $p < .05$ (one-tail)

---

[a] N = 10. A five-point scale was used in the first study.

[b] The residual scores and the ratings were converted to normalized standard scores.

correlations across ten teachers were perhaps an artifact of two quite "deviant" teachers.

A more conservative test was made by transforming the variable and residual scores to T-scores, normalized scores with mean of 50 and stnadard deviation of 10, and then recalculating the correlations. These recalculated results are also presented in Table 1. Even with this test, the ratings on following the format, corrections, requiring 100% criterion responding, and signals were very accurate predictors of student achievement.

Discussion

Perhaps the most important aspect of this first study was obtaining a very significant correlation between teacher behavior and student achievement in a highly prescribed program such as Distar. Such variation in teacher behavior and student achievement (adjusted for prior knowledge) suggests that even a highly specified program such as Distar cannot be considered a single variable. Although the teacher is provided with a script to follow word for word, no two classrooms are receiving the same instruction (although probably more so in the Distar program than in any other program). Rather, there is a good deal of variation in teacher and student behavior. Thir confirms the importance of studying what kinds of variations produce optimum gains on measures of interest.

Although the predictive importance of several variables have been tentatively demonstrated, the meaning of the correlations is not completely clear. Is a given variable important throughout the program or does the effectiveness of the particular behavior vary from week to week or perhaps from concept to concept? Secondly, teachers were rated and compared on different lessons. This could be a problem, since it is possible that one lesson would not be as difficult to teach

as another--thus favoring the teacher who is teaching a less difficult lesson. These are new researchable hypotheses which must be answered before we can begin to understand how the teacher affects pupil achievement in the Distar program.

Study Two

The purpose of the second study was to refine the methodological problems that existed in the first study. The chosen variables were examined in the context of a short-term (13 lessons or days) study with only one concept being recorded for analysis, rather than all the concepts being taught in the program.

Method

Materials. The multiple attributes sequence was chosen because it is probably the most difficult skill taught in the Distar Language I program. Thus the teachers would necessarily have to demonstrate maximum use of their teaching skills acquired during training.

The purpose of the introductory lessons in multiple attributes is to teach the child that all of the characteristics in a descriptive statement must be true of the object described for the statement to be valid. An example of a lesson from the program is presented in Figure 2. Notice that in order for the child to say, "This dog is little and wearing a hat," the dog must be both little and wearing a hat. If the dog is either not little, or not wearing a hat, or not little and not wearing a hat, the correct statement describing each of the three situations is, "This dog is not (little and wearing a hat)." The reason is different in each case, but the negation (not) can precede the multiple attributes in each of the three cases.

Procedure. Eight teachers teaching three groups each were selected. Each child in the group, following lesson 90 and prior to lesson 91 (the first day of multiple attributes instruction) was individually given an 18-item verbally administered criterion referenced pretest on multiple attributes. The test was constructed so as to ask the child to:

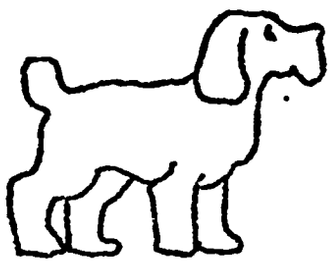
Praise the children for correct responses. Correct mistakes immediately.

Task 2

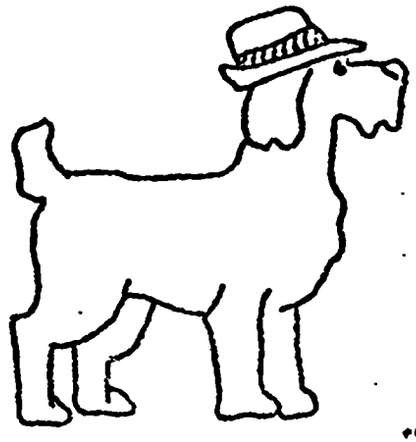
Group Activity

a. Point to the dogs. What are these? THESE ARE DOGS. Call on one child.
Find the dogs that are little. Everybody, tell me about these dogs.
Repeat, using the dogs that are wearing a hat, are big, and are not wearing a hat.

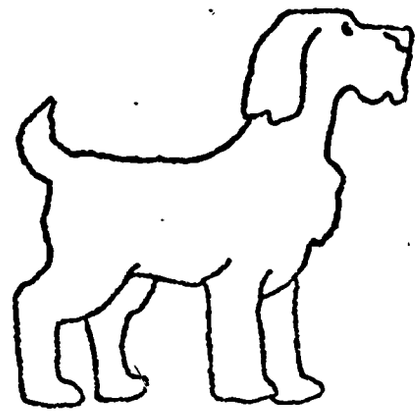
b.
Point. Is this dog little? YES.
Is this dog wearing a hat? YES.
Is this dog little and wearing a hat? YES.
Say the whole thing.
THIS DOG IS LITTLE AND WEARING A HAT.



c.
Point. Is this dog little? YES.
Is this dog wearing a hat? NO.
Is this dog little and wearing a hat? NO.
Say the whole thing.
THIS DOG IS NOT LITTLE AND WEARING A HAT. Why not?
THIS DOG IS NOT WEARING A HAT.



d.
Point. Is this dog little? NO.
Is this dog wearing a hat? YES.
Is this dog little and wearing a hat? NO.
Say the whole thing.
THIS DOG IS NOT LITTLE AND WEARING A HAT. Why not?
THIS DOG IS NOT LITTLE.



e.
Point. Is this dog little? NO.
Is this dog wearing a hat? NO.
Is this dog little and wearing a hat? NO.
Say the whole thing.
THIS DOG IS NOT LITTLE AND WEARING A HAT. Why not?
THIS DOG IS NOT LITTLE AND NOT WEARING A HAT.

Individual Activity

f. Call on individual children. Point to a picture. Is this dog little and wearing a hat? If the dog is not little and wearing a hat, ask: Why not?

Figure Example of a multiple attributes lesson. What the teacher says is underscored. How the children are to respond is capitalized.

1. Answer questions by indicating whether identifying features are present in an object and to make full statements describing two identifying features of an object.

2. Distinguish how a picture differs from a statement about that picture by making complete statements that describe the object pictured.

As in the first study, none of the test items was similar to the examples used in the lessons, except in form. Four lessons (93, 96, 100, and 102) were chosen for analysis and each teacher was audiotaped while teaching the lesson.

Observation Instrument. Because all teachers taught identical lessons, it was possible for the rating scale to be very specific for five of the categories. A seven-point scale was used for these categories. The last two categories (praise and feedback and pacing within tasks) were less specific and rated on the same five-point scale used in the first study.

The categories and the criteria for receiving a particular rating were as follows:

1. Follows the Format--Group Tasks

   7: Says all the words in the teacher's script. Does not omit words and does not add words except to praise, correct, or require 100% criterion responding.

   6: Says nearly all the words, deleting or adding incidental words which do not change the signals or intent of the format.

   5: Less than 6 but not omitting major portions of the task.

   4: Leaves out complete statement or signal from the major focus of the format. May lead responses.

   3: Adds or deletes words that change the major focus of the format.

   2: Deletes major portions of the format.

   1: Ignores the teacher's script completely.

2. Follows the Format--Individual Tasks

Identical to the 7-point scale for Follows the Format--Group Tasks.

3. Corrections

7: Corrects all mistakes as they occur according to the following paradigm (this paradigm is modified from the complete correction paradigm because the mistakes made during multiple attributes instruction are almost always a lack of information):

a. Give the answer;
b. Test the child(ren) on the segment of the task missed;
c. Repeat the entire task from the beginning.

6: Corrects all mistakes as they occur by giving the answer and testing the child(ren) on the segment of the task missed,
Does not repeat the entire task from the beginning.

5: Provides the answer most of the time.
Either tests the child(ren) on the segment of the task missed,
OR
Repeats the entire task from the beginning but not both.

4: Usually does not provide the answer.
Either tests the child(rem) on the segment of the task missed,
OR
Repeats the entire task from the beginning but not both.

3: Provides the answer most of the time.
Does not test the child(ren) nor repeat the entire task from the beginning.

2: Rarely provides the answer.
Does not test the child(ren) nor repeat the entire task from the beginning.

1: Ignores all mistakes as they occur.

4. Requires 100% Criterion Responding

A rating of 7-4 is characterized by the teacher allowing the children to respond again after having made a mistake.

7: Always holds the group to the signals until children respond correctly.
Goes back and repeats entire task until correct; follows correction paradigm.

6: Follows correction paradigm.
Repeats task but perhaps not until every child responds correctly--but nearly 100%.

5: When a mistake is made, usually gives the answer and tests segments.
Does not repeat entire task.
May recycle through portions of the task to firm-up responding.

4: Repeats entire task or tests only.
Does not provide answer when a child lacks information.
May recycle through portions of a task to firm-up responding.

A rating of 3-1 is characterized by the teacher not allowing the children to respond again after having made a mistake.

3: Usually provides the answer when a child lacks information.
Does not usually test segments missed.
Does not usually repeat the entire task.
Rarely recycles through protions of the task.

2: Rarely provides the answer when a child lacks information.
Does not test segments missed.
Does not repeat the entire task or portions of the task.

1: Does not provide the answer when a child lacks information.
Does not test segments missed.
Does not repeat the entire task or portions of the task.

5. Signals--Multiple Attributes

7: Pauses before all attributes.
Treats attributes as a unit by running the attributes together as if they were one word.
Does not emphasize "and."

6: Pauses before all attributes.
Treats attributes as a unit.
May emphasize "and" but does not require the children to emphasize "and."

5: Pauses before most attributes.
Does not usually treat attributes as a unit.

4: Does not pause before attributes.
Treats attributes as a unit.

3: Either pauses before attributes and emphasizes "and" and requires children to emphasize "and"
OR

Does not usually pause before the attributes and does not treat the attributes as a unit.

2: Pauses rarely before attributes.
Does not treat attributes as a unit; emphasizes "and"
Requires children to emphasize "and."

1: Wrong signal is presented.
Never pauses.
Never treats attributes as a unit.

6. Praise and Feedback (See Study 1)

7. Pacing Within Tasks (See Study 1)

Two graduate students rated all recordings while following the exact script the teacher was to follow for that particular lesson. An observation code was developed so that a written record could be kept of the transactions.

Results

In Table 2 the mean scores and standard deviations for the pretest, the seven categories, and the posttest are presented. Correlations between the ratings on the seven categories and student achievement on the multiple-attributes test (adjusted by regression for the mean pre-test score) are also presented. Ratings on the variables of following the format (group and individual tasks), corrections, requiring 100% criterion responding, and signals were most predictive of student achievement. Inter-rater reliability ranged from .88 to .98.

Discussion

The second study represented a semi-replication of the first study. Although eight of the ten teachers in the first study were used in the second study, a different part of the language program was chosen for analysis and an expanded and more specific rating system was developed.

Table 2

Correlations for Study 2[a]

	Means	Standard Deviations	Part Correlations	T-Score Part[b] Correlations	Partial Correlations
Pretest Score	6.70	1.83			
Posttest Score	14.82	2.46			
1. Follows the Format Group Tasks	5.75	0.67	.44*	.45*	.49**
2. Follows the Format Individual Tasks	5.21	1.41	.43*	.41*	.43*
3. Corrections	4.12	1.00	.44*	.39*	.45*
4. Requires 100% Criterion Responding	4.53	0.93	.60***	.61***	.52***
5. Signals[c]	4.16	0.98	.67***	.69***	.66***
6. Praise and Feedback	3.26	0.68	.42*	.33	.41*
7. Pacing Within Tasks	3.01	0.49	.26	.20	.25

*** $p < .005$ (one-tail)

** $p < .01$ (one-tail)

* $p < .05$ (one-tail)

---

[a] N = 24. A seven-point scale was used in the second study for categories 1 - 5. A five-point scale was used for categories 6 - 7.

[b] The residual scores and the ratings were converted to normalized standard scores.

[c] The meaning of this category differs for each study.

Interestingly, the most critical variable in the second study was "signals." This suggests that when teaching multiple attributes the teacher should pause before the attributes and treat them as a unit. (This is of course conjecture and must be experimentally studied before we can assert that pausing before the attributes and treating them as a unit enhances learning.)

Furthermore, when the children and teacher "say the whole thing" with a not statement (see the picture of the little dog not wearing a hat in the figure on page 20), the results of the study suggest that it perhaps is important to pause after the "not:" This dog is not.....little and wearing a hat. In this way the children can "see" that the not refers to the unit as a whole. Typically what happens in the above example when the teacher does not pause is that the children interpret the teacher as saying that the dog is not little and that the dog is wearing a hat (the exact opposite is true). The results of the second study certainly point to this problem in presenting the multiple attributes format.

Perhaps the reason "corrections" diminished in predictive ability is that when the teacher corrected the children's mistakes, the teacher did not pause in the appropriate points in the format--thus leaving the children confused after "correcting!" It thus appears that differences in results as compared to the first study reflect the different types of tasks taught in the two studies. This suggests that the relative importance of instructional skills possibly varies according to the learning task.

It is also conceivable that the importance of instructional behaviors varies according to the lesson in the program. For example, a teacher who is superior in requiring 100% criterion responding during the first lessons of the program would be teaching the children a content-independent concept (Engelmann, 1969a.) about working in school. That is, "the teacher is requiring me to respond correctly to every signal 100% of the time." After a teacher has established this concept, she can lessen her

behavior of requiring of requiring 100% criterion responding (less reminders to the children that they have to respond together when the signal is presented, etc.) because the children's behavior indicates that they have learned this concept--that they respond to every signal, in unison.

To test the hypothesis that the importance of instructional behavior varies across the program, it would be necessary to sample teacher behavior on an interval schedule and concurrently test the children on the material covered and pretest them on the next interval. For example, the children would be pretested on lessons 1 through 15; systematic observations taken on the teacher's behavior during this period; a posttest given at lesson 15 and another pretest given on lessons 16 through 30; systematic observations taken; and so forth. Correlations (adjusted for prior knowledge) would be computed to determine the relative relationship between teacher behavior and student cognitive gains at each interval.

Conclusions

The major importance of these studies is in the validation (so far) of a methodology. The approach was to study the relationship between student outcomes such as achievement and those teacher behaviors that were most emphasized in the training sessions and teacher's manuals. The methodology was tested within the context of a highly structured curriculum program--the Distar System, a direct instructional approach to teaching beginning reading, arithmetic, and language.

Within the Distar Language program, four variables remained in predictive importance across the two studies: 1) the extent to which the teacher follows the lesson format; 2) the ability to correct according to a prespecified paradigm all mistakes as they occur; 3) the degree to which the teacher requires 100% criterion responding to each signal; and 4) the extent to which the teacher pauses before presenting a clear signal and requires unison responses. The second study furthermore suggested that the importance of certain instructional behaviors may vary throughout the program.

The predictive importance of the above four variables was replicated by recoding the audiotapes from the first study using a modified observation instrument and obtaining a parsimony of description with principal components analysis (Siegel, et al., 1972). The major conclusions of that analysis were:

1. In a predictive sense it is not only important to attempt to correct mistkaes when they occur but it is also important to correct the mistakes according to the correction paradigm.

2. In a predictive sense it is important that the teacher get unison responses from the group. That is, none of the children should be allowed to cue off of other children's responses.

3. In a predictive sense, praise (general and specific) for appropriate responding and attending behavior and feedback (repeating the correct answer) is unimportant. This of course does not mean that it is unimportant for things other than achievement, for example, humaneness or civility or positive self-image.

4. In a predictive sense it is important to follow the format--both for group and individual tasks. Slight modifications or improvements in the format are permissible. (p. 27)

Perhaps the most interesting result was to obtain a very significant correlation between teacher behavior and student achievement in a highly structured program such as Distar. This suggests that even in a curriculum program that controls teacher behavior to the extent that it specifies word for word what to say to a group of students, there remains a large amount of variation in both teacher behavior and student performance. As Gallagher (1966) has concluded in his study of BSCS teachers:

> The data would suggest that there really is no such thing as a BSCS curriculum presentation in the schools. . . each teacher filters the materials through his own perceptions, and to say that a student has been through the BSCS curriculum probably does not give as much specific information as the curriculum innovators might have hoped. (p. 33)

This underscores the importance of studying the kinds of variation within a curriculum materials package that produce desired changes in student behavior.

List of References

Baker, E. L. Relationship between learner achievement and instructional principles stressed during teacher preparation. Journal of Educational Research, 1969, 63, 99-102.

Bissel, J. S. Implementation of planned variation in head start. I. Review and summary of the Stanford research institute interim report: first year of evaluation. Office of Child Development, U.S. Department of Health, Education, and Welfare, 1971.

Engelmann, S. Conceptual Learning. San Rafael: Dimensions Publishing Co., 1969 (a).

Engelmann, S. Preventing Failure in the Primary Grades. Chicago: Science Research Associates, 1969 (b).

Engelmann, S., & Bruner, E. Distar reading I. Chicago: Science Research Associates, 1969.

Engelmann, S., & Bruner, E. Distar reading II. Chicago: Science Research Associates, 1970.

Engelmann, S., & Carnine, D. Distar arithmetic I. Chicago: Science Research Associates, 1969.

Engelmann, S., & Carnine, D. Distar arithmetic II. Chicago: Science Research Associates, 1970.

Engelmann, S., & Carnine, D. Distar arithmetic III. Chicago: Science Research Associates, 1972.

Engelmann, S., Osborn, J., & Engelmann, T. Distar language I. Chicago: Science Research Associates, 1969.

Engelmann, S., Osborn, J. Distar language II. Chicago: Science Research Associates, 1970.

Engelmann, S., Osborn, J. Distar language III. Chicago: Science Research Associates, 1972.

Engelmann, S., & Stern, S. Distar reading III. Chicago: Science Research Associates, 1972.

Flanders, N. A. Analyzing teaching behavior. Reading, Massachusetts: Addison-Wesley, 1970.

Gage, N. L. Teaching methods. In R. L. Ebel (Ed.), Encyclopedia of Educational Research (4th Ed.). London: Macmillan, 1969, 1446-1458.

Gallagher, J. J. Teacher variation in concept presentation of the BSCS curriculum program. Urbana: Institute for Research on Exceptionsl Children, University of Illinois, 1966.

Gallagher, J. J. Analyses of teacher classroom strategies associated with student cognitive and affective performance. Cooperative Research Project No. 3325. Urbana: University of Illinois, 1966.

Katz, L. G. Children and teachers in two types of head start classes. Young Children, 1969, 24, 342-349.

Kochendorfer, L. H. The development of a student checklist to determine classroom teaching practices in high school biology; Classroom practices of high school biology teachers using different curriculum materials; and Use of the biology classroom activity checklist in identifying specific classroom practices of individual teachers and students. In A. E. Lee (Ed.), Research and curriculum development in science education: I. The new programs in high school biology. University of Texas at Austin, 1967, 71-78; 79-84; 85-89.

La Shier, W. S., Jr. The use of interaction analyses in BSCS laboratory block classrooms. Journal of Teacher Education, 1967, 18, 439-446.

Lindvall, C. M., & Cox, R. C. The IPI evaluation program: AERA evaluation monograph no. 5. Chicago: Rand Mc Nally, 1970.

Maccoby, E. E., & Zellner, M. Experiments in primary education: Aspects of project Follow-Through. New York: Harcourt Brace Jovanovich, Inc., 1970.

Medley, D. M., & Mitzel, H. E. Measuring classroom behavior by systematc observation. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Niedermeyer, F. C., & Dalrymple, M. C. The development and evaluation of an exportable teacher training package for a kindergarten reading curriculum. Paper presented to the meeting of the American Educational Research Association, February 1970.

Olivero, J. L. Developing the oral language program. Albuquerque, New Mexico: Southwestern Cooperative Educational Laboratory, undated.

Rosenshine, B. Evaluation of classroom instruction. Review of Educational Research, 40, 279-301, 1970.

Rosenshine, B. New directions for research on teaching. Paper presented at the conference on How Teachers Make a Difference, sponsored by the Bureau of Education Personnel Development, U.S. Office of Education, Washington, D. C., April, 1971.

Rosenshine, B., & Furst, N. Research on teacher performance criteria. In B. O. Smit (Ed.) Research on teacher education: A symposium. Engelwood Cliffs, New Jersey: Prentice-Hall, 1971, 37-72.

Rosenshine, B., & Furst, N. The study of teaching in natural settings using direct observation. In Travers, R. M. W. (Ed.), Second Handbook of Research on Teaching. Chicago: Rand McNally, 1973.

Siegel, M. A., Rosenshine, B., Marzano, W., & Walsh, P. A multivariate comparison of two program-specific observation instruments. Unpublished manuscript, Bureau of Educational Research, University of Illinois, 1972.

Simon, A. & Boyer, E. G. (Eds.) Mirrors for behavior: an anthology of classroom observation instruments. Vols. 1-14 and Summary and Supplementary Vols. A and B. Philadelphia: Research for better schools, 1967, 1970a, 1970b.

Soar, R. S. Advantages of multiple systems over a single system. Paper presented to the American Educational Research Association, 1971. Gainesville: Institute for Development of Human Resources, College of Education, University of Florida.

Soar, R. S., Soar, R. M., & Ragosta, M. The validation of an observation system for classroom management. Paper presented at the meeting of the American Educational Research Association, February 1971, Gainesville: University of Florida.

Tatsuoka, M. M. Nationwide evaluation and experimental design. Paper presented at the meeting of the American Educational Research Association, April 1972, University of Illinois.

Walberg, H. J. Predicting class learning: an approach to the class as a social system. American Education Research Journal, 1969, 6, 529-542.